A Novel Malware Detection Framework Based on Innate Immunity and Danger Theory

Mohamed Ahmed Mohamed Ali and Mohd Aizaini Maarof

Abstract Artificial immune system (AIS) is a computational system inspired by the principles and processes of the Biological immune system which has the capabilities to learn, adapt, self tolerance and memories actions, which make it a good example that we can take for solving some major problems in many fields, including the problem of malware detection in the field of computer security. The main idea is to detect any type of files that trying to harm the computer system by infecting some executable software when these files running, spread it to other files or computers. In this paper, we proposed a framework to detect malware using the innate immune system combined with danger theory to eliminate tow major drawbacks of current malware detection methods; detection accuracy and high false positive alarms.

Keywords Innate immune system · Danger theory · Malware detection

1 Introduction

The main obstacles facing the traditional malware detection methods were the high rate of creating new malware, the ability to change their shapes from time to time and from place to place (polymorphic malware) which make the detection use the

M. A. M. Ali (✉)
Faculty of Mathematical Sciences, University of Khartoum, Khartoum, Sudan
e-mail: mamautm@gmail.com;mama@uofk.edu

M. A. Maarof
Faculty of Computer Science and Information Systems, Universiti Teknology Malaysia,
UTM Skudai 81310 Johor, Malaysia
e-mail: aizaini@utm.my

K. J. Kim and K.-Y. Chung (eds.), *IT Convergence and Security 2012*,
Lecture Notes in Electrical Engineering 215, DOI: 10.1007/978-94-007-5860-5_4,

normal model for detecting malware based on the saved data (Signature-base model) a useless job [1]. However, in the last two decades the field of the artificial immune system (AIS) creates a new research area help the researchers to overcome efficiently some problems in the field of computer science like pattern recognition, data mining, intrusion detection and malware detection [2]. The biological immune system (BIS) is a system of biological structures and processes within an organism that protects against disease by identifying and killing pathogens and tumor cells. It detects a wide variety of agents, from viruses to parasitic worms through the integration of its two parts, innate and adaptive. It needs to distinguish pathogens from the organism's own healthy cells and tissues in order to function properly [3, 4]. Detection is complicated as pathogens can evolve rapidly; producing adaptations that avoid the immune system and allow the pathogens to successfully infect their hosts, but with the main characteristics of the biological immune system like: adaptability, self- tolerance, diversity, distributable and saved memory make it easier to defeat any invaders were trying to harm the organism [5]. Artificial immune system (AIS) inherits these characteristics to overcome many problems in the field of computer security. In section two we introduce the concept of innate immune system, section three discuss the danger theory concept and its benefits, then we introduce the novel framework in section four.

2 Innate Immune System

The innate immune system represents the first line of defense in the human immune system containing some external parts like skin, mucous and stomach acids to keep pathogens out of the body and internal parts like the inflammatory response and phagocytes. Phagocytes are a class of cells (part of the white blood cells) can engulf pathogens through its surface receptors which had the ability to connect to the proteins on the pathogen surface. After this connection is happen the phagocytes cut the bacteria or virus protein into small parts called peptide to attach them to major histocompatibility complex type 2 (MHC II) to present this complex on the phagocytes surface. Phagocytes called antigen presenting cells (APCs) when they present the complex of MHC II and the peptide on its surface. Phagocytes cells comprise macrophages, neutrophils and dendritic cells. Macrophages and neutrophils are phagocytes (cellular engulfment) the invading pathogen, then killing them through a respiratory burst. The neutrophils are the numerically superior cells of white blood cells (WBC) and the faster one to receive the infected tissue. Dendritic cells include the basophils and the eosinophils, although they are categorized as phagocytic cells, they are not killing the pathogen by phagocytosis. The basophils mediate the allergic reaction, while the eosinophils kill the invader pathogen by secreting highly toxic proteins added to the Toll like receptors (TLR) which are a pathogen recognition receptors found in the cell membrane also activate the immune cell response [5–7].

3 Danger Theory

The concept of danger theory initialized by Matzinger disprove that the immune system defense mechanism depend on the definition of what is part of the organism cells and what is not, what we called (self nonself theory SNS) which treat any things coming from outside as an invader [8]. Danger theory declares that interaction of the B cell receptors with antigen initiates a signal, this signal initiate the immune response. B cells secrete specific antibodies that recognize and react to stimuli. Another type of cell, the T (killer) lymphocyte, is also important in different types of immune reactions. The Danger model added another layer of cells and signals proposing that antigen presenting cells (APCs) are activated by danger alarm signals from injured cells, such as those exposed to pathogens, toxins, and so forth [9]. Alarm signal scanned to be constitutive or inducible, intracellular or secreted, or even a part of the extracellular matrix. Because cells dying by normal programmed processes are usually scavenged before they disintegrate, whereas cells that die necrotically release their contents, any intracellular product could potentially be a danger signal when released. Inducible alarm signals could include any substance made, or modified, by distressed or injured cells [10].

4 A Novel Malware Detection Framework Based on (Innate-Danger)

The proposed framework in Fig. 1 is composed of three main phases: the preprocessing phase, the innate immunity phase, and the danger theory phase.

4.1 Phase 1: Preprocessing

The preprocessing phase objectives is to define the self from non self depend on the data collected as a training data set to select the features that is suitable for classifying the malware from benign files and create an adaptive memory to store all the information about known malware as a signature based detection. The feature selection process means that specific features that are affected by the malware and for different types of attacks should be given the higher priority in selection compared to the features that keep out of change. By the end we will get the most important features that will help in differentiating between malware and benign files. By the end of this phase, we get the features and information about self and non self.

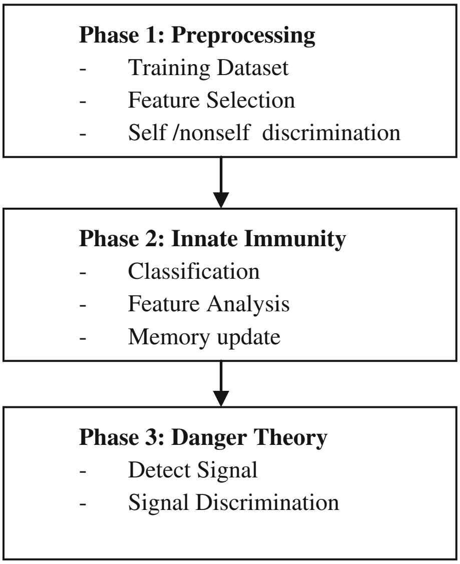

Fig. 1 The proposed framework

4.2 Phase 2: Innate Immunity

After complete the preprocessing phase we go through the innate immunity phase where we make the classification based on the features selected in the previous stage. These features help us to take advantages of the innate characteristics to distinguish between the harmful file and the benign file. Fetching the data in this phase leading to two processes, the first process is making a decision if the file is a malware or not depend on the signature (Pathogen Associated Molecular Patterns-PAMPs) associated with that file in hand [11]. If it is a malware then we stimulate the system to increase the detection ability of that type of malicious software and update the memory. The second process taking place if it is not a known malware we then look at the features selected from the previous stage whether the file have some of the malware features or not, if it had some of these features then maybe this is a malware and maybe not, but we must to make sure before take a decision to avoid the false positive alarms if the file is not a malware. In this case we move forward to the danger theory phase.

4.3 Phase 3: Danger Theory

In this phase we take the data from the previous phase to avoid the problem that facing the most of the accurate detection models which is the high rate of the positive alarms as a result of classifying wrongly some benign files as malware due to the lack of information about the file being scanned. In that case we increase the percentage of the accuracy with time tradeoffs spending in dealing with wrong alarms. So we add this phase to decrease the false positive alarm. By taking the output of the innate phase we come with some files have a number of malware features. By applying the maturation examination to the signals and executables file we can eliminate the false alarms and make the detection status not positive until the triggering of signal happened. Selection of the signal depends on the features like processor (CPU) usage and memory usage or any other feature. We here select the file that make a high memory usage as a danger signal.

5 Conclusions

Malware detection is a major task nowadays not only because the importance of the information and the resources store and transfer and process these information, also because the big evolution in the creation of malware and the related malware detection industry. A lot of models and frameworks proposed during the last two decades, but have their limitations because the accuracy and false positive alarms tradeoffs. In this paper a novel malware detection framework based on innate immunity and danger theory to overcome the low detection accuracy and the high rate of false positive alarms. This work is exploratory in which many experiments will be conducted to verify the viability of the framework.

Acknowledgments This work and research is done by support of Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Malaysia, Faculty of Mathematical Sciences, University of Khartoum, Sudan.

References

1. Christodorescu M, Jha S, Seshia SA, Song D, Bryant RE (2005) Semantics-aware malware detection. In: IEEE symposium on security and privacy, 2005
2. Castro LND, Von Zuben FJ (1999) Artificial immune systems: part I—basic theory and applications. Technical Report, RT–DCA 01/99, Dec 1999
3. Timmis J, Knight T, Castro LND, Hart E (2004) An overview of artificial immune systems. 2004
4. Andrews L (2008) Immunity, St. Martin's Minotaur 2008
5. Kuby J (1994) Immunology. vol 2nd edn. 1994
6. Parkin J, Cohen B (2001) An overview of the immune system. The Lancet 357(9270):1777–1789
7. Medzhitov R (2001) Toll-like receptors and innate immunity. Nat Rev Immunol 1(2):135–145
8. Matzinger P (1994) Tolerance, danger, and the extended family. Annu Rev Immunol 12:991–1045

9. Ali MAM, Maarof MA (2012) Malware detection techniques using artificial immune system. In: Kim KJ, Ahn SJ, (eds) Proceedings of the international conference on IT convergence and security 2011, Springer, Netherlands, pp 575–587
10. Matzinger P (2002) The danger model: a renewed sense of self. Science 296(5566):301–305
11. Janeway CA (1989) Approaching the asymptote? Evolution and revolution in immunology. Cold spring harbor symposia on quantitative biology, vol 54 Pt 1, pp 1–13